Economics 20900: Introduction to Econometrics: Honors

Department of Economics
University of Chicago
Spring 2007

Lecturer: Sam Schulhofer-Wohl, sschulh1@uchicago.edu
Teaching assistant: Andy Zuppann, zuppann@uchicago.edu

All course materials will be posted on the class Web page at chalk.uchicago.edu.

1 Class times and office hours

Lectures: 10:30 to 11:50 a.m. Tuesdays and Thursdays in Rosenwald 015.

Problem sessions: 5 to 6 p.m. Mondays in Rosenwald 015.

Sam's office hours: 7 to 8 p.m. Tuesdays on A-level of Regenstein Library, or by appointment.

Andy's office hours: 2 to 4 p.m. Fridays on A-level of Regenstein Library, or by appointment.

2 Course description

This course is an introduction to empirical methods in economics. We will discuss how economists combine statistical tools and economic theory to answer questions about the world.

Example: People who serve in the military receive a great deal of training. Does this training allow veterans to earn higher wages than non-veterans? Can we answer the question simply by comparing the wages of some randomly chosen veterans and non-veterans?

Objectives: By the end of the course, you should be able to perform basic empirical analyses; critically evaluate and explain basic econometric work done by other people; and analyze the theoretical properties of estimators commonly used in basic empirical work.

Prerequisites: This course is intended for students who are contemplating graduate study in economics and have substantial previous background in mathematics and statistics. Statistics 24400 or the equivalent would be ideal. Compared with Economics 21000 (Econometrics A), this course puts more emphasis on abstract statistical theory and less emphasis on applications. The last page of this syllabus is a math and statistics pre-test. If the material on the pre-test is unfamiliar to you, or if you want to focus mainly on applications, you should consider taking Econometrics A instead of this course.

3 Grading and policies

Final exam – 10 a.m. to noon Friday, June 1. *Note that this is not the time listed in the registrar's official final exam schedule.* Closed book, closed notes.

Midterm exam – Tuesday, May 1, in class. Closed book, closed notes.

Problem sets – About seven, due most Thursdays at the beginning of class. The purpose of the problem sets is almost entirely to give you practice in doing econometrics, not to evaluate your progress. We will grade one randomly chosen problem on each problem set and will drop your two lowest grades. We will also distribute detailed solutions, which you should make sure to review and understand. You will need to use the software program Stata, which is available in the public computer labs. You may work with classmates but must write up your own answers. *Late problem sets will not be accepted.*

Replication exercises – Due at the beginning of class Tuesday, April 24, and Tuesday, May 29. One week before each due date, I will distribute a published econometrics paper from a professional journal and the raw data used in that paper. You must then use Stata to attempt to replicate all of the empirical results in the paper. Your write-up should be about five pages and should consist of 1) a critical summary of the paper, describing the economic question it asks, the data and methods used to answer this question, and your views on whether the results are credible; 2) a description of your efforts to reproduce the empirical results, including an explanation of any difficulties encountered; and 3) the Stata program you used to try to reproduce the results. Replication code for the papers may already be available on the Web; if so, you are *not* permitted to look at or use it. You may consult with classmates but must ultimately write your own programs and your own report. *Except in case of a documented medical or family emergency, late assignments will lose one letter grade per day or fraction thereof that they are late.*

Grading – Because this is an honors course, I will not grade on a curve. I encourage you to help each other learn the material. I will weight the items according to whichever of the following schemes is better for you:

- Final 40%, midterm 30%, problem sets 10%, replication exercises 20%.
- Final 65%, problem sets 10%, replication exercises 25%.

This scheme effectively makes the midterm optional (although I very strongly encourage you to take it); hence, there will be no make-up exams for students who miss the midterm for any reason.

If you require accommodations for a disability, please see me as soon as possible.

Any student who cheats on an exam or engages in any other form of academic dishonesty will fail the course and be reported to the dean.

4 Readings

There is no perfect advanced econometrics book. We will use two main texts that have complementary strengths:

- Hansen, Bruce, 2007, *Econometrics*, PDF available for free on Chalk and at the author's Web site: `www.ssc.wisc.edu/~bhansen/notes/notes.htm`.
- Ruud, Paul A., 2000, *An Introduction to Classical Econometric Theory*, Oxford University Press.

You are responsible for material in Hansen and Ruud only if we also cover it in class or on an assignment. If you find Hansen and Ruud unhelpful, some other good books are:

- Davidson, Russell, and James G. MacKinnon, 1993, *Estimation and inference in econometrics*, Oxford University Press. Slightly easier than other graduate-level books.
- Goldberger, Arthur S., 1991, *A Course in Econometrics*, Harvard University Press. A classic older graduate-level book.
- Greene, William H., 2003, *Econometric Analysis*, Prentice Hall. The standard reference and a good "cookbook" of techniques.
- Hayashi, Fumio, 2000, *Econometrics*, Princeton University Press. Emphasizes GMM as a unifying framework.
- Wooldridge, Jeffrey M., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press. Typically used for second-year graduate courses, but the first four chapters are a good review of basic asymptotic theory.
- Wooldridge, Jeffrey M., 2005, *Introductory Econometrics: A Modern Approach*, Thomson South-Western. A good introductory book, and the text for Econ 21000.

I have placed all of these books on reserve at Regenstein Library.

5 Course outline

1. Statistical background.
   a. Unconditional and conditional distributions; conditional densities, means and variances. Hansen, chapter 2 and appendix B; Ruud, appendix D.
   b. Modes of convergence: convergence in probability, convergence in distribution. Laws of large numbers, central limit theorems. Hansen, appendix C.
   c. Estimation theory: method of moments, maximum likelihood. Hansen, appendix D; Ruud, appendix E.

2. Ordinary least squares regression.

   a. Mechanics. Hansen, sections 3.1-3.5. (See also the linear algebra review in Hansen, appendix A, and Ruud, appendix C.)

   b. Assumptions of the classical linear regression model. Sampling distribution of the OLS estimators. Hansen, sections 3.8 and section 4.11; Ruud, chapters 6, 8 and 10.

   c. Interpretations of OLS: projection, method of moments, maximum likelihood. Ruud, chapters 2 and 3; Hansen, sections 3.6 and 3.7.

   d. Modeling: collinearity, dummy variables, logs, polynomials, transformations. Ruud, sections 1.1, 2.1, 3.1 and 5.2.

3. Asymptotic inference.

   a. Consistency and asymptotic normality. Hansen, sections 4.1-4.3; Ruud, chapter 10.

   b. Efficiency. Gauss-Markov theorem. Hansen, sections 3.9 and 3.10; Ruud, chapter 9.

   c. Homoskedasticity and heteroskedasticity. Estimating variance-covariance matrices. Hansen, sections 4.4 and 4.5; Ruud, sections 18.1, 18.2 and 18.4.

   d. Weighted least squares and generalized least squares. Hansen, section 5.1; Ruud, section 18.5.

   e. T-tests, F-tests and confidence regions. Hansen, sections 4.6-4.10; Ruud, chapter 11.

   f. Classes of tests: Wald, likelihood ratio, Lagrange multiplier. Ruud, chapter 17.

4. Endogeneity, simultaneous equations and instrumental variables.

   a. Irrelevant variables and omitted variables. Hansen, sections 5.8 and 5.9; Ruud, sections 20.1-20.3.

   b. Measurement error. Hansen, introduction to Chapter 9; Ruud, examples in chapter 20.

   c. Reduced form and structural equations. Identification. Hansen, sections 9.2 and 9.3; Ruud, examples in chapter 20.

   d. Two-stage least squares. Hansen, sections 9.4 and 9.5; Ruud, sections 20.5 and 20.6.

   e. Generalized method of moments. Hansen, chapter 7; Ruud, chapters 21 and 22.

Math and statistics pre-test

If the questions below seem very difficult, you may prefer to take Econometrics A instead of this course.

1. What is the difference between a sample average and a population mean of a random variable?

2. Let X be a random variable with density $f(x) = (1/\sqrt{2\pi\sigma^2}) \exp\left[-(x-\mu)^2/2\sigma^2\right]$. What is the name of this distribution? What is the mean of X? What is the variance of X? Show how to calculate the mean and variance of X using the definitions of mean and variance.

3. Let Z be a random variable with density $f(z) = k \exp(-\theta z)$ for $z \geq 0$ and some real number k. What is the name of this distribution? What must be the value of k? What is the CDF of Z?

4. Suppose you have N data points: $x_1, x_2, \ldots, x_N$. Give formulas for the sample average and sample standard deviation of these data. Also, let $\bar{x}$ represent the sample average and prove that
$$\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \bar{x}^2.$$

5. Suppose you measure the heights of a random sample of 100 male and 100 female University of Chicago students. Among the men in your sample, the mean height is 69 inches and the standard deviation is 2 inches. Among the women, the mean height is 64 inches and the standard deviation is 1 inch. Test each of the following null hypotheses against an appropriate alternative: 1) the average height of male students is 70 inches; 2) the average height of female students is 60 inches; 3) the average heights of male and female students are the same.

6. What is a confidence interval?

7. Consider the following matrices:
$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 9 & 8 \\ 7 & 6 \\ 5 & 4 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1.5 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1.5 & 2 \\ 3 & 5 \end{bmatrix}$$
For each matrix: What is its rank? What is its column space? Does it have an inverse? If the inverse exists, calculate it. Also, calculate each of the following or say why the calculation is impossible: $\mathbf{AB}$, $\mathbf{BA}$, $\mathbf{CB}$, $\mathbf{C'B}$, $\mathbf{B'C}$, $\mathbf{C} + \mathbf{D}$, $\mathbf{D} - \mathbf{A}$.

8. Calculate the partial derivatives of $f(x, y, \gamma, \theta) = (y - x\theta)^\gamma$ with respect to x, y, γ, θ.